ALPHY/AIEM 2022 – submitted abstracts

==========

1 Sophie Abby sophie.abby@univ-grenoble-alpes.fr
TIMC, CNRS, Université Grenoble Alpes

Title: Evolutionary scenario of substrate regio-selectivity for hydroxylases involved in ubiquinone production

Authors: Katayoun Kazemzadeh, Clothilde Chenal, Mahmoud Hajj Chehade, William Schmitt, Qiqi He, Manon Jarzynka, Nelle Varoquaux, Ludovic Pelosi, Ivan Junier, Fabien Pierrel, Sophie Abby

Abstract: The ubiquinone (UQ) is a lipophilic molecule which is essential to respiratory chains, as it is involved in shuttling electrons between respiratory enzymes. The UQ biosynthetic pathway includes three hydroxylation reactions on three positions of the UQ precursor: C5, C1 and C6. In Escherichia coli, these three hydroxylation steps are performed by three different enzymes whereas some bacteria use two enzymes (e.g. UbiL and Coq7 for Rhodospirillum rubrum), or even a single enzyme (e.g. UbiM for Nesseiria meningitidis) to perform the three reactions [1]. This denotes a variation in 'regio-selectivity' across UQ hydroxylases. Interestingly, the six known UQ hydroxylases belong to only two distinct protein families. Coq7 is a member of the di-iron carboxylate family, whereas all other hydroxylases identified so far (UbiF, -H, -I, -L, -M, thereafter called 'Ubi-FMO') belong to the large Flavin Mono-oxygenase family. The Ubi-FMO family offers the unique opportunity to study the evolution of enzymatic regio-selectivity among homologous enzymes. In this study, we combined experimental and computational approaches to investigate the evolution of regio-selectivity across UQ hydroxylases. Based on genomic distribution, we could predict expected regio-selectivity for the different UQ hydroxylases. We then performed phylogenetic analyses to (i) sample taxonomically representative sequences for experimental characterization, and (ii) investigate potential evolutionary scenarios for Ubi-FMO and the overall genomic repertoire in UQ hydroxylases. In order to test the enzymes' predicted activity, experimental characterizations of UQ hydroxylases were performed on an unprecedented scale for >100 hydroxylases across Proteobacteria using heterologous expressions in relevant KO mutants of E. coli. We could systematically confirm the anticipated regio-selectivity, showing that regio-selectivity is evolutionarily conserved for each different protein sub-family. We propose an evolutionary scenario of the evolution of UQ hydroxylases' regio-selectivity. Overall, this study provides an interesting case study for the evolution of enzymatic function, and paves the way for future research on the precise molecular mechanisms dictating enzymes' regio-selectivity.

[1] Pelosi L., Ducluzeau A. L., Loiseau L., Barras F., Schneider D., Junier I., Pierrel F. (2016) Evolution of Ubiquinone Biosynthesis: Multiple Proteobacterial Enzymes with Various Regioselectivities To Catalyze Three Contiguous Aromatic Hydroxylation Reactions. mSystems 1: e00091-16 DOI: 10.1128/mSystems.00091-16

==========

2 Juliette Archambeau juli.archambeau@orange.fr
University of Bordeaux (BIOGECO lab - INRAE)

Title: Extreme climatic events but not environmental heterogeneity shape within-population genetic variation in maritime pine

Authors: Juliette Archambeau, Marta Benito Garzon, Marina de Miguel Vega, Benjamin Brachi, Frédéric Barraquand and Santiago C. Gonzalez-Martìnez

Abstract: How evolutionary forces interact to maintain quantitative genetic variation within populations has been a matter of extensive theoretical debates. While mutation and migration increase genetic variation, natural selection and genetic drift are expected to deplete it. To date, levels of genetic variation observed in natural populations are hard to predict without accounting for other processes, such as balancing selection in heterogeneous environments. We aimed to empirically test three hypotheses: (i) admixed populations have higher quantitative genetic variation due to introgression from other gene pools, (ii) quantitative genetic variation is lower in populations from harsher environments (i.e. experiencing stronger selection), and (iii) quantitative genetic variation is higher in populations from spatially heterogeneous environments. We used phenotypic measurements of five growth, phenological and functional traits from three clonal common gardens, consisting of 523 clones from 33 populations of maritime pine (Pinus pinaster Aiton). Populations from harsher climates (mainly colder areas) showed lower genetic variation for height in the three common gardens thus suggesting a negative effect of climate-driven selection in these populations. This result was further validated on fully independent height data from an additional set of experiments. The robustness of our results was also supported by the lack of association between population admixture and genetic variation, suggesting no influence of gene flow on the genetic variation of the traits considered. Unexpectedly, however, we did not find any association between within-population genetic variation and environmental heterogeneity. Our study adds to the debate on the maintenance of genetic variation within populations by providing empirical support for the role of selection in reducing genetic variation in populations of a long-lived forest tree, and therefore indirectly their adaptive potential.

==========

3 Matteo Bisardi matteo.bisardi@gmail.com
ENS, Sorbonne Université

Title: Data-driven modeling of experimental protein evolution

Authors: Matteo Bisardi, Juan Rodriguez-Rivas, Francesco Zamponi and Martin Weigt

Abstract: Thanks to the explosion of protein sequence data available, enabled by next generation sequencing, unsupervised machine learning methods can be exploited to computationally infer complex protein fitness landscapes. In particular, methods like Direct Coupling Analysis (DCA), which take into account conservation and coevolution between sites, have proven successful in recovering with high accuracy residue-residue contacts, in predicting the phenotypic effect of mutations, and in artificially generating in vivo functional sequences. We discuss here a recent application of DCA in the context of experimental molecular evolution. We show that a simple stochastic dynamics in such data-driven fitness landscapes can quantitatively predict important evolutionary signatures of experimentally evolved libraries. Fitness distributions, position-specific mutational profiles and the emergence of epistasis are well reproduced; our modeling framework offers also a quantitative explanation for the different outcomes of recently published experiments. Finally, we discuss how this approach could be complemented by experimental data to design optimized protocols for protein evolution.

==========

4 Simon Boitard simon.boitard@inrae.fr
INRAE

Title: Heterogeneity in effective size across the genome: effects on the Inverse Instantaneous Coalescence Rate (IICR) and implications for demographic inference under linked selection

Authors: Simon Boitard, Armando Arredondo, Camille Noûs, Lounès Chikhi, Olivier Mazet

Abstract: The relative contribution of selection and neutrality in shaping species genetic diversity is one of the most central and controversial questions in evolutionary theory. Genomic data provide growing evidence that linked selection, i.e. the modification of genetic diversity at neutral sites through

linkage with selected sites, might be pervasive over the genome. Several studies proposed that linked selection could be modelled as first approximation by a local reduction (e.g. purifying selection, selective sweeps) or increase (e.g. balancing selection) of effective population size (Ne). At the genome-wide scale, this leads to variations of Ne from one region to another, reflecting the heterogeneity of selective constraints and recombination rates between regions. We investigate here the consequences of such genomic variations of Ne on the genome-wide distribution of coalescence times. The underlying motivation concerns the impact of linked selection on demographic inference, because the distribution of coalescence times is at the heart of several important demographic inference approaches. Using the concept of Inverse Instantaneous Coalescence Rate, we demonstrate that in a panmictic population, linked selection always results in a spurious apparent decrease of Ne along time. Balancing selection has a particularly large effect, even when it concerns a very small part of the genome. We also study more general models including genuine population size changes, population structure or transient selection and find that the effect of linked selection can be significantly reduced by that of population structure. The models and conclusions presented here are also relevant to the study of other biological processes generating apparent variations of Ne along the genome.

==========

5 Raphaël Bricout bricout@biologie.ens.fr
Institut de Biologie de l'ENS (IBENS), Département de biologie, Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
Title: Evolution is surprisingly non uniform along coding sequences
Authors: Raphaël Bricout, Auguste Genovesio, Hugues Roest Crollius

Abstract: An open question in the field of molecular evolution is to understand why codon or amino acid residues display such a range of evolutionary rates within protein sequences. Providing a mechanistic clue for this site heterogeneity would help building better models of protein evolution which are critical for phylogenetic inferences and to identify sites of adaptive selection. To explore this question, we rescaled all gene sequences to the same length and computed the dN/dS ratio per position to evaluate the mean changes of selective pressure along coding sequences. Aggregating data that belong to thousands of plant and animal gene families, we reveal a pattern that was concealed until now, where the dN/dS is on average much higher on protein edges compared to the middle of the sequences. Controls show that no significant bias in algorithms or data processing could explain our observation, suggesting a biological explanation. We show that predicted functional domains are depleted near protein edges, where amino-acid changes are much more frequent, and that these regions also correlate with increased solvent accessibility in the folded proteins. This finding is consistent with relaxed functional and structural constraints at protein edges, creating an environment that tolerates amino-acid changes with little consequences on protein function. We finally show that as a consequence, current methods designed to identify positive selection are biased towards protein edges, suggesting a confounding effect between positive selection and weak negative selection. Together our results provide a possible explanation for rate heterogeneities along coding sequences and a practical avenue to account for these in models of sequence evolution.

==========

6 Ewen Burban ewen.burban@univ-rennes1.fr
Université Rennes 1
Title: Gene network simulations provide testable predictions for the molecular domestication syndrome
Authors: Ewen Burban, Maud I Tenaillon, Arnaud Le Rouzic

Abstract: The domestication of plant species lead to repeatable morphological evolution, often referred to as the phenotypic domestication syndrome. Domestication is also associated with

important genomic changes, such as the loss of genetic diversity compared to adequately large wild populations, and modifications of gene expression patterns. Here, we explored theoretically the effect of a domestication-like scenario on the evolution of gene regulatory networks. We ran population genetics simulations in which individuals were featured by their genotype (an interaction matrix encoding a gene regulatory network) and their gene expressions, representing the phenotypic level. Our domestication scenario included a population bottleneck and a selection switch mimicking human-mediated directional and canalizing selection, i.e., change in the optimal gene expression level and selection towards more stable expression across environments. We showed that domestication profoundly alters genetic architectures. Based on four examples of plant domestication scenarios, our simulations predict (i) a drop in neutral allelic diversity, (ii) a change in gene expression variance that depends upon the domestication scenario, (iii) transient maladaptive plasticity, (iv) a deep rewiring of the gene regulatory networks, with a trend towards gain of regulatory interactions, and (v) a global increase in the genetic correlations among gene expressions, with a loss of modularity in the resulting coexpression patterns and in the underlying networks. We provide empirically testable predictions on the differences of genetic architectures between wild and domesticated forms. The characterization of such systematic evolutionary changes in the genetic architecture of traits contributes to define a molecular domestication syndrome.

==========

7 Thibaut Capblancq thibaut.capblancq@univ-grenoble-alpes.fr
Laboratoire d'Ecologie Alpine (LECA)

Title: Genomic prediction of population maladaptation to future climate: concept, methods and challenges

Authors: Thibaut Capblancq

Abstract: The current global climatic crisis has begun affecting the entire biosphere, posing a serious threat to the health and persistence of climate-sensitive populations and species. At the same time, recent technological advancements give us access to massive quantities of data pertinent to biodiversity conservation (e.g., genome-scale DNA sequencing, high-resolution climate models) and new sophisticated computational tools that can take advantage of these data to identify conservation risks and opportunities under a changing climate. A key question is how to harness the predictive power of statistical algorithms to generate actionable predictions about which populations will be most heavily impacted by climate change, and how these losses could be mitigated. Addressing this problem requires recognizing that not all populations within a species will experience climate change in the same way, nor will they respond uniformly to a given change in climate. Populations frequently display 'local adaptation' in which natural selection has modified their genetic composition to optimize traits and maximize their survival and growth in the local climate. As a consequence, populations may become vulnerable to changing climates that would disturb this optimized local adaptation. Scientists have been developing computational approaches that pair large genomic datasets with high-resolution climate forecasts to estimate the gene-climate relationship and identify populations most vulnerable under a changing climate. To do this, statistical models are used to assess association between the current genetic composition of populations and their climate of origin. Then, knowing the current gene-climate relationship, we use the models to project genetic variation under a different climate. Finally, we derive the expected shift in the gene-climate relationship between current and projected climates, allowing us to make quantitative predictions about the degree of expected maladaptation -- a metric that is also named 'genomic offset' or 'genomic vulnerability'. During this talk, I will describe the concept of genomic offset/vulnerability and explore methods and challenges associated with the use of this metric to predict the impacts of climate change on species and populations.

==========

8 Vincent Castric vincent.castric@univ-lille.fr
CNRS - université de Lille

Title: Evolution of the gene regulatory network controlling dominance/recessivity interactions between self-incompatibility alleles in Capsella and Arabidopsis
Authors: Jorg Bachmann, Marco Fracassetti, Andrew Tedder, Mathieu Genete, Eléonore Durand, Julie Ferreira de Carvalho, Vincent Castric, Tanja Slotte

Abstract: The level of functional constraint on genetic elements acting as dominance modifiers has been a hot topic in evolutionary genetics throughout the 20th century. The first ´†bona fide†ª dominance modifiers have been recently discovered in plants of the Brassicaceae family, where they govern dominance relationships between self-incompatibility alleles (S-alleles). These modifiers take the form of a simple regulatory circuit based of small RNA transcriptional regulators produced by dominant S-alleles and their target sites on recessive S-alleles. In this study, we take advantage of the largely shared repertoire of S-alleles between *Capsella grandiflora* and *Arabidopsis halleri* to evaluate the functional constraint on their network of dominance relationships across the approximately 8 million years since they diverged. We determine the extent to which the underlying regulatory circuit has been rewired, specifically testing whether the same sRNAs and target sites are being used, or whether new dominance modifiers have emerged. We also examine whether the regulatory interactions have been modified by the accumulation of mutations at the sRNAs or/or their targets, and whether they represent compensatory changes.

==========
9 Sophie-Carole Chobert sophie-carole.chobert@univ-grenoble-alpes.fr
TIMC CNRS
Title: Deciphering the distribution of quinone biosynthetic pathways across Proteobacteria
Authors: Sophie-Carole Chobert, Safa Berraies, Ludovic Pelosi, Ivan Junier, Fabien Pierrel, Sophie Abby

Abstract: Isoprenoid quinones are molecules that have a major role in bioenergetics as they shuttle electrons across the respiratory chains of most living organisms. There are different quinone types that can be discriminated by their mid-point redox potential. This potential determines with which respiratory enzymes the quinones function, the respiratory substrates that are accessible, and ultimately the environment in which the organisms can live. In Proteobacteria, the two main quinones are menaquinone (MK), a low potenial quinone (~-80 mV) and ubiquinone (UQ), a high potential one (~+100 mV). So far, UQ was considered well adapted for the respiration of dioxygen, while MK would rather be involved in anaerobic respiration in O2-deprived contexts. However, our team recently discovered an O2-independent biosynthetic pathway for UQ production, while the classical pathway depends on the presence of O2 [1]. It was shown that this O2-independent pathway is crucial for anaerobic respiration (denitrification) in Pseudomonas aeruginosa [2]. Thus, these discoveries challenge the respective assumed physiological roles and origins of the MK and UQ pathways in Proteobacteria. In this study, we systematically investigated the quinone production potential across Proteobacteria through the annotation of quinone (MK and UQ) biosynthetic pathways in a large set of complete genomes. Particular attention was paid to the genes specific of the UQ O2-independent pathway (ubiT, -U and -V) for which we started to reconstitute the evolutionary history. We also addressed the question of their genetic architecture and regulation. Altogether, this large-scale study gives us more insights while inviting us to revisit the classical view of the respective physiological roles of the respiratory quinones found in Proteobacteria.

[1] Pelosi L, Vo C-D-T, Abby SS, Loiseau L, Rascalou B, Chehade MH, Faivre B, Goussé M, Chenal C, Touati N, et al (2019) Ubiquinone Biosynthesis over the Entire O2 Range: Characterization of a Conserved O2-Independent Pathway. 10: 21
[2] Vo C-D-T, Michaud J, Elsen S, Faivre B, Bouveret E, Barras F, Fontecave M, Pierrel F, Lombard M & Pelosi L (2020) The O2-independent pathway of ubiquinone biosynthesis is essential for denitrification in Pseudomonas aeruginosa. Journal of Biological Chemistry 295: 9021-9032

==========

10 Yves Clément yves.clement@ijm.fr
Université de Paris - Institut Jacques Monod

Title: The effect of chromosomal rearrangements on base composition evolution in mammals

Authors: Yves Clément, Céline Bon, Hugues Roest Crollius

Abstract: Mammalian species harbor differences in GC content distributions, from a homogeneous distribution in mouse to a heterogeneous distribution in human. In mammalian species, GC content evolution is mainly affected by GC-biased gene conversion (gBGC), a neutral process that favors the fixation of G & C alleles in highly recombining regions. This process behaves like natural selection, and various studies have shown that effective population size is linked with gBGC efficiency and GC content evolution. Moreover, because this process is associated with the distribution of meiotic recombination and larger chromosomes harbour lower recombination rate than smaller chromosomes, the distribution and intensity of meiotic recombination within a genome is driven by chromosome size. As a consequence, chromosomal rearrangements affect GC content evolution: fusions or fissions changing chromosomal length affect recombination rates, while changing gene locations within a genome alter their recombination landscape. The precise role of chromosomal rearrangements in the diversity of GC content distribution in mammals is still an open question. We reconstructed ancestral GC content at 3rd codon positions (GC3) in almost 3000 genes in ancestral nodes of a phylogeny of 30 Boreoeutheria species using non-homogeneous models of evolution and used high precision ancestral karyotype reconstructions in Boreoeutheria species to link karyotype evolution and GC content evolution. We found that changes in gene positions along chromosomes determine how their GC3 evolves, and we explored the link between karyotype evolution and GC content evolution. We further investigated the homogenization of base composition in murid genomes. In murid, we pinpointed this shift in base composition to the Myomorpha-Muroidea lineage. By studying patterns of GC3 evolution in this lineage and in muntjak deers (a species that underwent massive recent chromosomal fusions), we found that this homogenization is not compatible with massive chromosome fusions and homogenization of chromosome length, but with massive rates of chromosomal rearrangements. Our results shed new light on the evolution of base composition, and the effects of chromosomal rearrangements on the distribution of GC content.

==========

11 Frédéric Delsuc Frederic.Delsuc@umontpellier.fr
CNRS Institut des Sciences de l'Evolution de Montpellier

Title: Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals

Authors: Rémi Allio*, Sophie Teullet*, Dave Lutgen*, Amandine Magdeleine, Rachid Koual, Marie-Ka Tilak, Benoit de Thoisy, Christopher A. Emerling, Tristan Lefébure, and Frédéric Delsuc

Abstract: Ant-eating mammals represent a textbook example of convergent morphological evolution. Among them, anteaters and pangolins exhibit the most extreme convergent phenotypes with complete tooth loss, elongated skulls, protrusive tongues, and powerful claws to rip open ant and termite nests. These two placental mammal lineages also possess hypertrophied salivary glands producing large quantities of saliva to capture and potentially digest their social insect prey. Despite this remarkable convergence, comparative genomic analyses have shown that anteaters and pangolins differ in their chitinase gene (CHIA) repertoires, which potentially degrade the chitinous exoskeletons of ingested ants and termites. While the lesser anteater (*Tamandua tetradactyla*) harbours four functional CHIA paralogs (CHIA1, CHIA2, CHIA3, and CHIA4), Asian pangolins (Manis spp.) have only one functional paralog (CHIA5). A recent transcriptomic analysis has shown that CHIA5 is highly expressed in all major digestive organs (stomach, pancreas, large intestine, and liver) of the Malayan pangolin (*Manis javanica*), including its tongue and salivary glands. Here, we present the first comparative transcriptomic analysis of salivary glands in 23 species of placental mammals, including

new ant-eating species and close relatives, together with complementary RNA-seq data for the major digestive organs of the lesser anteater. Our results on digestive enzyme gene expression show that salivary glands play a major role in adapting to a myrmecophagous diet. A detailed analysis of nine paralogous chitinase genes revealed that convergently-evolved pangolins and anteaters express different chitinases in their hypertrophied salivary glands and other digestive organs. CHIA5 is overexpressed in Malayan pangolin salivary glands and other digestive organs, whereas the lesser anteater exhibits high levels of CHIA3 and CHIA4 expression in salivary glands and other digestive organs. Overall, our results demonstrate that divergent molecular mechanisms underlie convergent adaptation to the ant-eating diet in pangolins and anteaters, highlighting the role of historical contingency and molecular tinkering of the chitin-digestive enzyme toolkit in this classical example of convergent evolution.

==========

12 Olivier Dennler olivier.dennler@irisa.fr
Université Rennes 1
Title: Evolution of motifs signature in ADAMTS / ADAMTS-like, correlate with PPI signature
Authors: Olivier Dennler, Samuel Blanquart, François Coste, Catherine Belleannée, Nathalie Théret

Abstract: ADAMTS and ADAMTS-like proteins are involved in microenvironment remodelling and are now considered as new potential therapeutic targets in numerous diseases. However, the characterization of these proteins is still under progress. Indeed, new approaches for their functional annotation need to be developed. Three main features of these proteins bring a challenge to conventional approaches for protein functional characterization. Namely, they are the numerous genes in the ADAMTS and ADAMTS-like family, the multi-domain structures of their encoded proteins and the limited knowledge available. We propose here a new method adapted to multi-paralog and multi-domain protein sequences based on module decomposition and phylogenetic inferences. The aim is to identify functional protein regions. This method combines and integrates multiple phylogenetic inference strategies ; 1) gene phylogeny inference, 2) ancestral modules composition inference using Domain-Gene-Species (DGS) reconciliation and 3) inference of function evolution. This new strategy is implemented as a pipeline allowing to correlate the conserved sequence modules and the functions evolution in order to identify the co-appearance of conserved modules and functions. Applying this method to ADAMTS-TSL proteins permits us to validate a known module and to propose new functional modules.

==========

13 Thomas Forest thomas.forest@college-de-france.fr
Collège de France
Title: POSTER : Quantify species decline based on genomic and occurrence data
Authors: Thomas Forest

Abstract: The quantification of demographic change is a common, yet tricky question in population genetics. Here, techniques based on genetic information (allele frequency spectra, recombination information) are experimented on a diverse set of animal taxa, which present different types of demographic structure and history at different time scales (decline, steady, growth): butterflies (*Zygaena corsica*), bumblebees (*Bombus confusus*), mussels (*Margaritifera auricularia*), birds (*Hirundo rustica, Picus viridis*), gorilla (*Gorilla gorilla*), etc. Some progress on species using genetic data will be presented : *Zygaena corsica*, a species that appears to be two different ones, and *Hirundo rustica*, a species under recent decline. The Site Frequency Spectrum, SFS, is a summary statistic which has many purposes, and our study suggests that SFS-based demographic inference techniques are indeed quite powerful. However, they seem more efficient for quantifying ancient decline rather than recent one using genome wide polymorphism data, suggesting that recombination-based techniques or combined approaches with occurrence data could improve our estimations for recent demographic changes.

==========

14 Jasmine Gamblin jasmine.gamblin@college-de-france.fr
Collège de France/Sorbonne Université
Title: Bottlenecks can constrain evolutionary paths
Authors: Jasmine Gamblin, François Blanquart, Amaury Lambert

Abstract: (POSTER) Population bottlenecks are commonplace in experimental evolution, specifically in serial passaging (or transfer) experiments. These consist in periodically subsampling a microbial population and placing it on a new medium to grow again. Serial transfer experiments are thought to reproduce in vitro natural living conditions of microbial species. However, the way such controlled demography influences the population genetics of these species remains unclear. Using a modeling approach and large limit techniques, we study the establishment and survival of two types of beneficial mutations with a trade-off between mutation rate and fitness advantage. We show that demographic parameters such as the initial population size and the time between two bottlenecks determine which paths may be taken by the evolving population, and also affect the timing of evolution.

==========

15 CarlosAlberto Gandarilla-Pérez carlosgandarilla88@gmail.com
Sorbonne Universités, UPMC Université Paris 06, CNRS, Biologie Computationnelle et Quantitative, Institut de Biologie Paris Seine, 75005 Paris, France.
Title: Graph alignment for protein similarity networks
Authors: Carlos A. Gandarilla-Perez, Anne-Florence Bitbol and Martin Weigt

Abstract: Specific protein-protein interactions are crucial in almost all complex biological processes, they play a key role in the formation and stability of multiprotein complexes. Functional interactions between proteins result in coevolution between the interaction partners, causing their sequences to be correlated. Recently, statistical inference methods applied on sequence data alone are able to exploit these correlations to address two conceptually related inference tasks: finding pairs of interacting proteins, and identifying pairs of residues which form contacts between interacting proteins. Algorithms based on Direct Coupling Analysis (DCA) and Mutual Information (MI) successfully predict partners among paralogs, but, they require large joint alignments of homologous protein pairs. Here, we exploit phylogenetic correlations coming from the similar evolutionary histories of both interacting proteins. We introduce an algorithm to align the similarity graphs of two interacting protein families involved in bacterial signal transduction, and to predict specific interaction partners between these families whose members are known to interact. We apply the algorithm to histidine kinases and response regulators from bacterial two-component signaling systems, and to proteins from ATP-binding cassette (ABC) transporter complexes. We obtain accurate predictions on these systems without using any a priori knowledge of interaction partners. Two models of similarity networks were aligned using Monte Carlo techniques. Results depend on the model and the dataset used.

==========

16 Alice Genestier alice.genestier@univ-lyon1.fr
Laboratoire de biométrie et biologie évolutive (LBBE) - Université Claude Bernard Lyon 1
Title: The influence of genetic dosage on PRDM9-dependent evolutionary dynamics of meiotic recombination
Authors: Alice Genestier, Nicolas Lartillot and Laurent Duret

Abstract: Meiosis is an important step in the eukaryotic life cycle during which recombination and proper chromosome segregation takes place. In mammals, recombination is regulated by the Prdm9 gene. This gene, which possesses a double function (recruitment of the double strand break machinery and facilitation of the pairing of homologous chromosomes), induces an intra-genomic Red Queen

resulting from the opposition of two antagonistic forces : erosion of the recombination landscape by biased gene conversion and positive selection on Prdm9. This Red Queen was previously modeled, but without taking into account the role of PRDM9 as a pairing facilitator. Accordingly, I developed a mechanistic model taking into account the dual role of PRDM9. This modeling work gives important insights into the Red Queen mechanism, thus completing previous studies. In particular, it reveals that positive selection of new PRDM9 alleles is due to the reduced symmetrical binding caused by the loss of high affinity binding sites and, on the other hand, it demonstrates the influence of the genetic dosage of PRDM9 on the dynamics of the Red Queen, which can result in negative selection on new PRDM9 alleles entering the population.

==========

17 Jérémy Guez jeremy.guez@mnhn.fr
MNHN

Title: Simulating cultural and demographic processes interactions in human population genetics

Authors: Jérémy Guez, Bruno Toupance, Romain Laurent, Evelyne Heyer, Frederic Austerlitz, Flora Jay

Abstract: In many human populations, a positive correlation in progeny size has been shown between parents and children, a process called Cultural Transmission of Reproductive Success (CTRS). This process can result from various non-genetic causes such as social influence from parents on their children, the increase of children survival when aunts and uncles are present (allocare) and the transmission of fitness increasing resources. Here we study genome evolution through time under CTRS with the help of numerous summary statistics from the literature. We show that CTRS has a double impact on population genetics: (1) effective population size decreases when CTRS starts, mimicking a population contraction, and increases to its original value when it stops (2) coalescent trees are imbalanced as shown in previous studies. After a long time under CTRS, the effect of effective population size contraction vanishes, leaving the effect of trees imbalance only, which yields U-shaped Site Frequency Spectrum (SFS). Because of CTRS, impact on genetics of other processes such as demographic events are wiped out rapidly due to the lower effective population size. Once CTRS is over, an expansion-like effect happens due to effective population size going back to normal. This effect lasts as long as 2N generations, unlike imbalance of coalescent trees which only lasts a few dozens of generations once CTRS is over. We also show that as CTRS has an impact on SFS shape, it creates a bias in SFS-based demography inference. Considering that CTRS has been shown to happen in numerous human and animal populations around the world, one should be cautious when inferring population past processes from genomic data without including CTRS in the model.

==========

18 Julien JOSEPH julien.joseph@ens-lyon.fr
LBBE, CNRS

Title: Fine-scale recombination landscapes differ among Prdm9-lacking amniotes

Authors: Julien JOSEPH, Nicolas LARTILLOT, Laurent DURET

Abstract: Meiotic recombination plays a major role in evolution by maintaining genetic diversity and thereby enhancing the efficiency of selection. Interestingly, the few fine-scale recombination maps currently available in vertebrate taxa revealed very strong heterogeneities of recombination rate along chromosomes, at the kb-scale. However, the 'raison d'être' of these recombination hotspots remains unclear. So far, two types of hotspots have been reported. In humans and mice, recombination hotspots are determined by PRDM9, a sequence-specific DNA-binding protein, which recruits Spo11 to form double-stranded breaks (DSBs) and thus initiates recombination at its target sites. In these species, there is a deficit of recombination within genes that are highly expressed in the germline. Conversely, in taxa that have lost Prdm9, such as canids or birds, an increase in recombination rate is observed in the vicinity of promoters, in particular those that correspond to CpG islands (CGIs). Interestingly, in Prdm9-knockout mice, recombination is predominantly initiated at CGI-type

promoters. It has therefore been suggested that gene-promoter regions correspond to the 'default recombination hotspots' in absence of PRDM9, and that the function of PRDM9 is to target recombination away from these functional elements. In this study, we revisit this interpretation by investigating fine-scale recombination landscapes in dogs and chicken. In dogs, we confirm that CGIs correspond to recombination hotspots. We show that their recombination rate correlates negatively with their level of DNA methylation, consistent with patterns observed in Prdm9-KO mice. Conversely, we show that in chicken, CGIs do not correspond to recombination hotspots. The increase in recombination rate in the neighborhood of chicken CGIs is explained by a co-occurrence between CGIs and replication origins, with no link to germline methylation. Interestingly, recombination rates are negatively correlated to transcription levels in dogs, as in humans and mice, but positively correlated in chickens. This indicates that the associations reported between gene expression and recombination are not dependent on Prdm9. These observations highlight the diversity of fine-scale recombination landscapes in metazoan genomes, which remains largely unexplored.

==========

19 Caroline Kebaili caroline.kebaili@univ-grenoble-alpes.fr
Université Grenoble-Alpes
Title: Demographic inferences and climatic niche modelling shed light on the evolutionary history of the Apollo butterfly at regional scale
Authors: Caroline Kebaili, Stéphanie Sherpa, Delphine Rioux, Laurence Després

Abstract: Cold-adapted species escape climate warming by latitudinal and/or altitudinal range shifts, and currently occur in Southern Europe in isolated mountain ranges within 'sky islands'. We studied the genetic structure of the Apollo butterfly in five such sky islands (above 1,000 m) in France, and infer its demographic history since the last interglacial, using single nucleotide polymorphisms (ddRADseq SNPs). The Auvergne and Alps populations show strong genetic differentiation but not alpine massifs, although separated by deep valleys. Combining three complementary demographic inference methods and species distribution models (SDMs) we show that the LIG period was highly unfavourable for Apollo that probably survived in small populations in the highest summits of Auvergne. The population shifted downslope and expanded eastward between LIG and LGM throughout the large climatically suitable Rhône valley between the glaciated summits of Auvergne and Alps. The Auvergne and Alps populations started diverging before the LGM but remained largely connected till the mid-Holocene. Population decline in Auvergne was more gradual but started before (~7 kya vs. 800 ya), and was much stronger with current population size ten times lower than in the Alps. In the Alps, the low genetic structure and limited evidence for isolation by distance suggest a nonequilibrium metapopulation functioning. The core Apollo population experienced cycles of contraction- expansion with climate fluctuations with largely interconnected populations overtime according to a 'metapopulation-pulsar' functioning. This study demonstrates the power of combining demographic inferences and SDMs to determine past and future evolutionary trajectories of an endangered species at a regional scale.

==========

20 Joel Klein joel.klein@wur.nl
Wageningen University & Research
Title: Lineage specific loss of nodulation genes in the nitrogen fixation clade
Authors: Joel Klein & Rene Geurts

Abstract: Nitrogen-fixing root nodule endosymbiosis is a mutualistic interaction between some plant species and nitrogen-fixing bacteria. This so-called nodulation trait is present in species that make up ten taxonomic lineages, of which legumes are most well-known. The taxonomic lineages of nodulating plants are interspersed with lineages where nodulation is absent. Recent phylogenomic studies revealed that genes essential for nodulation are independently lost in non-nodulating relatives of

nitrogen-fixing plants. It led to the hypothesis that nodulation evolved only once and was lost multiple times. However, this single gain-multiple loss hypothesis is currently supported by the finding of only three genes that associate with the nodulation trait. To find additional support of this hypothesis, we aimed to identify lineage-specific loss of genes in non-nodulating species. To do so, we analyzed 80 genomes of plant species of the Fabid clade, which encompasses the nodulating lineages. We identified gene loss for 281 putative nodulation genes in at least one taxonomic lineage of non-nodulating plants. From this study, it became apparent that there is a significant variation in gene loss patterns in different lineages. We hypothesize that the differences in gene loss results from lineage-specific regulatory adaptation of nodulation genes. If nodulation genes are neo-functionalized to function exclusively in nodules, gene loss is likely in non-nodulating sister species.

==========

21 Ingrid Lafontaine ingrid.lafontaine@ibpc.fr
Sorbonne Université
Title:
Authors:

Abstract: Je peux faire un topo si vous manquez de candidats!

==========

22 Tanguy Lallemand tanguy.lallemand@inrae.fr
IRHS
Title: Chromosomal dominance in apple after Whole Genome Duplication
Authors: Tanguy Lallemand , Sébastien Aubourg , Jean-Marc Celton and Claudine Landès

Abstract: Polyploidy is an important driver of genetic evolution in plants genomes [1]. A whole genome duplication (WGD) dated around 50 million years ago [2] has been confirmed by obtaining a high quality genome for the domesticated apple (Malus domestica Borkh) [3]. This WGD is well conserved and allowed identification of chromosome segments with homologous genes in an conserved order called syntenic blocks. 63% of genes are in a syntenic block. In the context of WGD, homologous genes can be called ohnologous genes. A significant QTL imbalance between syntenic chromosome blocks has been identified by studying the projection of 589 QTL covering several phenotypic traits onto a single physical map. To understand the origin of this imbalance, we investigated whether the Transposable Elements (TE) environment of genes and expression differences between pairs of ohnologous genes could be associated with the observed QTL imbalance. A mapping pipeline of RNA-Seq datas was written to study potential transcriptional imbalance. It was used on all publicly available data sets of sufficient quality. Differential analyses were performed following this mapping. Classically differential analyses test genes in two conditions, here we tested differential expression between pairs of syntenic genes within the same experimental condition. In order to study TE, we wrote a pipeline computing TE density, TE coverage and number of TE in gene neighbourhood. Genes were then clustered using a unsupervised clustering method to classify them into 5 categories representing their TE density and coverage. Using data from 144 RNA-Seq experiments, our results show that out of the 16,376 ohnologous gene pairs, 10,376 genes are differentially expressed. Wilcoxon-Mann-Whitney tests were conducted on each syntenic chromosome pairs for each RNA-Seq experiments. P-values were aggregated using Fisher exact method allowing to identify 6 chromosome pairs significantly imbalanced, which is consistent with the observed QTL imbalance. The TE analysis suggests that for most ohnologous chromosome pairs, the distribution of genes within the different clusters is significantly different (chi-square test). Furthermore, in order to determine the direction of chromosome dominance, we tested each pair of ohnologous chromosoms (Wilcoxon-Mann-Whitney test). Several chromosome pairs are significantly unbalanced in terms of TE coverage. The direction of TE dominance is anti-correlated with the dominance of gene expression. Future analyses aim to

integrate these results using multinomial logistic regression and to explore GO enrichment of these genes unbalanced in expression and in TE environment.

[1] Pamela S Soltis and Douglas E Soltis. Current Opinion in Plant Biology, 30:159-165, 2016.
[2] Riccardo Velasco, Andrey Zharkikh, et al. Nat. Genet., 42(10):833-839.
[3] Celton, Daccord et al. Nat Genet, 49(7):1099-1106, 2017.

==========

23 Romuald Laso-Jadart romuald.laso-jadart@mnhn.fr
Museum National d'Histoire Naturelle
Title: Population genomics of the great white shark Carcharodon carcharias
Authors: Romuald Laso-Jadart, Elise Gay, Gavin Naylor, Stefano Mona

Abstract: Modelling the evolutionary history of a species requires a careful understanding of its population structure. Indeed, species are rarely (if ever) panmictic and the gene genealogy of a sample of lineages is affected by the migration rates connecting the array of sub-populations (hereafter, demes). In consequence, applying methods assuming panmixia can bias our interpretation of the variation of effective population size (Ne) through time, which is particularly worrisome when investigating species of conservation concern. In this study, we focused on the great white shark (Carcharodon carcharias), by whole-genome sequencing twelve individuals sampled in four locations (Western Indian Ocean, Southern and Northern Pacific Ocean, Norther Atlantic Ocean). The great white shark is considered vulnerable by the IUCN but is still characterized by large-scale migration potential: defining unit of management and understanding demographic trends is therefore crucial for its conservation. By applying a combination of inferential methods based on the sequential coalescent and the site frequency spectrum coupled with clustering algorithm, we propose a complex demographic scenario of ancient decline and recent fragmentation. Using the instantaneous inverse coalescence rate inferential framework, we interpret the decline in the ancestral Ne observed in all individuals/populations under panmictic models as the progressive decrease in the migration rates connecting a small number of demes, ultimately leading to fragmentation around 10 ky b.p.. We further observed a recent apparent increase in Ne which is therefore due to the establishment of a collecting phase of a stepping stone matrix where the South Africa-Australia corridor represents the connecting point between the North Pacific and the North Atlantic Oceans. Together, these results imply that the great white shark is presently highly threatened, and is probably unable to efficiently restore a proper genetic diversity through successful migratory events.

==========

24 Guillaume Laval glaval@pasteur.fr
Institut Pasteur
Title: Genome-wide estimation of natural selection parameters based on ancient DNA data
Authors: Gaspard Kerner, Guillaume Laval, and Lluis Quintana-Murci

Abstract: A number of studies in humans showed that ancient DNA data (aDNA) provide valuable information to formally estimate key natural selection parameters under various selection models, including positive and negative selection. Using 1,013 ancient European genomes covering multiple epochs over the last 10,000 years and an Approximate Bayesian Computation (ABC) framework we recently estimated the onset of selection (t) and the coefficient of negative selection (s) at the TYK2 P1104A polymorphism, a known monogenic predisposition to tuberculosis, highlighting the heavy burden on human health imposed by TB. Using more than 2,500 ancient genomes, we extended this work by conducting a genome-wide ABC estimation of these two selection parameters at 520,670 aDNA polymorphisms assuming realistic admixture scenarios and several models of natural selection. We first confirmed the ages and intensities of selection previously estimated for various iconic examples of positive selection in Europe, as those obtained for the putative selection targets in the

LCT (rs4988235, with t = 5,000 [3,000-9,000] years ago and s = 0.08 [0.065-0.095]) or in the TLR1,6,10 cluster regions (rs4321646, with t = 6,500 [4,500-10,000] years ago and s = 0.02 [0.01-0.06]). We also identified loci under recent polygenic adaptation in the context of increased risks to autoinflammatory disorders, with the particular contribution of variants at the SLC22A4, ALDH2 or FUT2 genes, all playing pleiotropic effects, as involved in different disorders. Finally, we found that the variants with the lowest negative selection coefficients are enriched in innate immunity genes, such as IL23R and TYK2 (mycobacterial infections), or LBP (bacterial infections).

==========

25 Polit Lélia lelia.polit@gmail.com
Sorbonne University

Title: Dietary adaptation in Neandertal, Denisovan and Sapiens revealed by gene copy number variation

Authors: Lélia Polit*, Riccardo Vicedomini*, Silvana Condemi, Laura Longo, Alessandra Carbone

Abstract: Dietary adaptation is the acquisition of an efficient system to digest food available in an ecosystem. The study of dietary adaptation has been mainly performed using the stable isotopic approach. We collected stable isotope data for more than 70 individuals, one of the largest pool for this kind of data ever used, and confirmed that we could not distinguish Neandertal from Sapiens behaviour. Hence, we looked at the genomic contribution to dietary adaptation. We searched 16 genomes from Neandertal, Denisovan and Early Sapiens and looked for food digestion genes that tend to have more or fewer copies than the modern human reference genome hg19. We identified 11 genes, including three gene clusters, with discernible copy number variation (CNV) trends at the population level. The genomic variation shows how metabolic pathways for lipid, brown fat, protein or carbohydrate metabolism adapt to metabolize food from animal or plant sources. Interpreting the copy number profiles of the 16 archaic individuals in relation to fossil evidence shows that Homo sapiens had an evolutionary advantage compared to Neandertal and Denisovan in adapting to cold and temperate ecosystems. In addition, we showed that both stable isotopes and CNVs are relevant features to distinguish Neandertal from Sapiens populations.

==========

26 Hugo Menet hugo.menet@univ-lyon1.fr
LBBE, Lyon 1

Title: Host-Symbiont-Gene phylogenetic reconciliation

Authors: Hugo Menet, Vincent Daubin, Eric Tannier

Abstract: Biological systems like holobionts are made of entities at many scales (macro-organisms, micro-organisms, genes...) which are, on one hand, bound to a common history because they all function together and depend on the others, and on another hand, driven by their individual interests. The evolution of such a system is approached by phylogenetic reconciliation, which describes the co-evolution of two different levels, genes and species, or hosts and symbionts for example. The limit to two levels has confined the use of reconciliation to either molecular studies on genes and species trees, or ecological studies on host-symbiont associations. Now, the holobiont concept is an occasion to gather all these scales by modeling multi-level inter-dependencies. We present a probabilistic model of evolution of three nested levels of organization that can account for the co-evolution of hosts, symbionts and their genes. This model allows gene transfer as well as host switch, gene duplication as well as symbiont diversification inside a host, gene or symbiont loss. Given three phylogenetic trees, we devise a Monte Carlo algorithm able to infer joint scenarios and compute their likelihood in the model, accounting for gene transfer rates dependence on host symbiont reconciliation as well as the impact of ghost lineages on these rates. On simulated dataset we compare our method to a gene symbiont reconciliation unaware of the host. We evaluate its capacity to infer horizontal gene transfer and show the possibility to differentiate simulations where horizontal gene transfer between

symbionts depend on their host and the cases where they do not. We show our results on a biological dataset consisting of 1034 gene families across 119 strains of Helicobacter pylori, a bacterial pathogen that is believed to have followed its human host during ancestral migrations, demonstrating how we can use our method to test different hypothesis of human population history.

==========

27 François Monnet monnet.fr.pro@gmail.com
Université de Lille
Title: Grey-zone of speciation in the green world
Authors: François Monnet, Xavier Vekemans, Yves Van De Peer, Camille Roux

Abstract: The speciation process consists in the progressive accumulation of 'reproductive barriers' (mutations that contribute to the reproductive isolation) between two populations, until a threshold is reached at which gene exchange is no longer possible. Although speciation genomics approaches have revealed the genomic location and molecular basis of such reproductive barriers in several study cases, there is still much to learn about the dynamics of this process. Considerable insight has been gained on the dynamics of speciation in animals from a recent meta-analysis of 61 species pairs. Among those, we learned that 1) genomic exchanges are widespread between distant species and 2) there is a 2% threshold of net divergence beyond which hybridization is generally suppressed. To date, we do not know whether this pattern is universal or specific to animals. Using the important amount of genomic data publicly available online nowadays, we draw a first picture of the relation between divergence and accumulation of reproductive barriers in the plant kingdom. To do so, we apply a tool based on the ABC (Approximate Bayesian Computation) approach which uses genomic data to infer the most probable scenario between different demographic patterns of isolation or migration between two populations. The results will also be used to study the impact of different life history traits on speciation.

==========

28 Lucas Paoli paolil@ethz.ch
Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Z¸rich
Title: Probing microbial evolutionary processes through population-level biogeography of the global ocean microbiome
Authors: Lucas Paoli, Hans-Joachim Ruscheweyh and Shinichi Sunagawa

Abstract: As natural microbial populations collectively drive global biogeochemical cycles and underpin human health, it is crucial to understand the processes that shape them. These eco-evolutionary processes can be studied by analysing biogeographical patterns, from local to global scale and across environmental gradients. Microbial biogeography has traditionally been investigated through taxonomic composition using ribosomal marker genes. However, the limited taxonomic resolution of these markers precludes detecting within-species biogeographical patterns that emerge from population-level evolutionary processes. The recent availability of global metagenomic surveys, notably in the case of the ocean microbiome, combined with reference genomes provides a unique opportunity to achieve the required resolution to study these mechanisms. Yet, such genome-resolved analyses have largely been restricted to cultivable microbes, which typically accounts for less than 10% of the global ocean microbiome genomic content. To overcome these limitations, we used large-scale bioinformatic binning to reconstruct >25,000 ocean microbial genomes from >1,000 metagenomes spanning global-scale surveys and time series. These efforts recovered >5,000 ocean microbial species, many of them from under-sampled phylogenetic lineages. Leveraging this genome collection in combination with dimension reduction and density-based clustering approaches, we identified discrete populations, often structured along environmental gradients, within most cosmopolitan marine microbial species. Our work demonstrates that much of the ocean microbiome population-level diversity remains underexplored, highlights the importance of evolutionary processes in shaping

its genomic makeup and lays the groundwork for a better understanding of these processes in natural microbial populations.

==========

29 Tom Parée paree@bio.ens.psl.eu
IBENS
Title: Reduced selective interference increases experimental adaptive rates in *C. elegans*
Authors: Parée Tom*, Noble Luke, Teotonio Henrique

Abstract: Theory suggests that by reducing selective interference recombination increases adaptive rates. Here, we use experimental evolution in Caenorhabditis elegans populations with alternative modifiers of the recombination rate landscape to test for adaptation. This nematode exhibits a wild-type recombination landscape with low recombination rates in chromosomal centers and high recombination rates chromosomal arms. Loss of function of the rec-1 gene equalizes recombination rates between centers and arms, without changes in the total recombination map length. Further, in our populations, most fitness loci are located in the chromosomal arms. We challenged C. elegans populations with standing genetic variation but different rec-1 alleles (wild-type vs. loss-of-function mutant) to a novel environment. After 40 generations of evolution, we measured a fitness-proxy and, during the experiment, patterns of genome-wide SNP allele frequency change. These data indicate that the wild-type rec-1 allele increases adaptive rates by decreasing the extent selection interference, thus confirming theoretical expectations.

==========

30 Flavia Pavan flavia.pavan.etu@univ-lille.fr
Unité Evolution Ecology Paleontologie
Title: Polymorphism and evolution of microRNA genes in *Arabidopsis halleri*
Authors: Flavia Pavan, Chloé Beaumont, Sophie Gallina, Clément Mazoyer, Mathieu Genete, Eléonore Durand, Firas Louis, Vincent Castric, Sylvain Legrand

Abstract: MicroRNAs (miRNAs) are a class of small non-coding RNAs that play important regulatory roles in plant and animal genomes. They are produced from a primary transcript forming a short foldback structure from which a mature miRNA is processed and negatively regulates a series of mRNA targets. Some miRNA genes are conserved over long evolutionary time scales, while others have apparently emerged more recently. To compare the functional constraint over the evolution of these two classes of miRNA genes, we performed preliminary polymorphism analysis of 72 *Arabidopsis halleri* individuals from across the species range. Our resequencing strategy was based on targeted sequence capture of 185 miRNA genes and their predicted target sites. We identified the homologous miRNAs in 88 plant species and determined the time since they emerged. We observed that the mature miRNA sequence presented lower nucleotide diversity than the other parts of the hairpin (stem, terminal loop). The miRNA-targeted region on the CDS was also less polymorphic than the neighboring regions. This suggests that evolution of the mature miRNA sequence and its complementarity target sequence are constrained by purifying selection. In addition, the evolutionarily conserved miRNA hairpins and their mature miRNA sequences presented less nucleotide diversity than the more recently emerged ones, suggesting that younger miRNA genes generally evolve under weaker selective constraint than older ones. These preliminary findings will be completed by the analysis of all miRNA genes in *A. halleri* and in her sister species *Arabidopsis lyrata*.

==========

31 Apolline Petit petit@egce.cnrs-gif.fr
EGCE
Title: How selection shapes gene regulatory network to enhance evolvability
Authors: Apolline Petit

Abstract: Can evolvability evolve to produce fitter mutants phenotypes ? Theoretically, the environment might have an impact on genetic architecture through time by selectioning architectures favoring fitter new mutant phenotypes. For a better insight of such a phenomenon, we studied the evolution of mutational correlations among gene expression levels in a gene regulatory network model. We tested multiple environments and initial gene networks to understand how likely this phenomenon is to occur in biological systems and how it is linked with gene network evolution. We showed that, in such a complex genetic architecture, mutational correlations could evolve as a response to the fitness function, but the evolution of co-expressions was constrained by the network topology. We found what interactions in gene regulatory networks impact mutational correlation and how it determines evolutionary pathway of mutational effects. In addition, we offer an effective way to quantify evolvability response to the fitness function.

==========

32 Pauline Raimondeau pauline.raimondeau@univ-tlse3.fr
Université Toulouse 3

Title: Biparental inheritance of plastids and intra-individual selection as a driver of parallel acceleration of genome evolution in two *Oleaceae* lineages?

Authors: Pauline Raimondeau, Céline Van de Paer, Hervé Phillipe, Pascal-Antoine Christin, Guillaume Besnard

Abstract: Rates of molecular evolution vary dramatically across genes and genomes. The multiple causes of this heterogeneity are difficult to disentangle. Organellar genomes of photosynthetic plants are thought to evolve under strong purifying selection and usually show little variation in evolutionary rates. Accelerated plastid evolution has been evidenced in several lineages, but the influence of selection has been generally disregarded. Here, we analyze the rates of plastome evolution among an extensive sampling of 117 Oleaceae. Plastid-specific acceleration of evolutionary rates was observed in two Oleaceae lineages. Comparison of the rates of synonymous and non-synonymous substitutions between the plastid and mitochondrial genomes revealed that this acceleration is mostly driven by non-synonymous changes. Using codon models, we showed that these patterns are driven by positive selection on a subset of non-photosynthetic genes. These shifts coincide with transitions to non-strict maternal inheritance of plastids, and we hypothesize that potential biparental transmission generates intracellular competition between undifferentiated plastids in zygotes, with strong selection for mutations that increase plastid proliferation leading to an excess of non-synonymous mutations. Our analyses show that selective processes can explain the acceleration of plastid evolution in some plant lineages.

==========

33 Lauric Reynes lauric.reynes@sb-roscoff.fr
IRL Evolutionary Biology and Ecology of Algae, CNRS, Sorbonne Université, Roscoff

Title: Genomic study of the connectivity of Mediterranean brown algae, the case of Ericaria zosteroides

Authors: Lauric Reynes, Didier Aurelle, Cristele Chevalier, Christel Pinazo, Myriam Valero, Stéphane Mauger , Aurélie Blanfuné, Charles-François Boudouresque, Marc Verlaque and Thierry Thibaut

Abstract: Dispersal is a central process that affects population growth, gene flow, and ultimately species persistence. Here we investigate the extent to which gene flow occurs between fragmented populations of the deep-water brown algae Ericaria zosteroides (Turner) Greville (Sargassaceae, Fucales). These investigations were performed at different spatial scales from the bay of Marseille (western Provence) to Corsica. As dispersal of zygotes is shown to be limited over distances beyond a few meters, we used a multidisciplinary approach, based on Lagrangian modeling and population genomics to test the hypothesis that drifting of fertile parts of thallus (eggs on fertile branches), mediated by ocean currents, enable occasional gene flow between populations. Therefore we

assessed the respective contribution of oceanographic connectivity, geographical isolation, and seawater temperatures to the genetic structure of this species. The genetic structure was assessed using 10 755 neutral SNPs and 12 outlier SNPs genotyped by dd-RAD sequencing in 261 individuals of E. zosteroides. We find that oceanographic connectivity is the best predictor of genetic structure, while differentiation in outlier SNPs can be explained by the depth of populations, as emphasized by the minimum seawater temperature predictor. However, further investigations will be necessary for clarifying how depth drives adaptive genetic differentiation in E. zosteroides. Our analyses revealed that local hydrodynamic conditions are correlated with the very high divergence of one population in the Bay of Marseille. Overall, the levels of gene flow mediated by drifting were certainly not sufficient to counteract differentiation by local genetic drift, but enough to allow colonization several kilometers away. This study stresses the need to consider secondary dispersal mechanisms of presumed low dispersal marine species to improve inference of population connectivity.

==========

34 Juan Rodriguez-Rivas juan.rodriguez.rivas@gmail.com
Sorbonne University
Title: Epistatic models predict mutable sites in SARS-CoV-2 proteins and epitopes
Authors: Juan Rodriguez-Rivas*, Giancarlo Croce*, Maureen Muscat, Martin Weigt

Abstract: The emergence of new variants of SARS-CoV-2 is a major concern given their potential impact on the transmissibility and pathogenicity of the virus as well as the efficacy of therapeutic interventions. Here, we predict the mutability of all positions in SARS-CoV-2 protein domains to forecast the appearance of unseen variants. Using sequence data from other coronaviruses, pre-existing to SARS-CoV-2, we build statistical models that do not only capture amino-acid conservation but more complex patterns resulting from epistasis. We show that these models are notably superior to conservation profiles in estimating the already observable SARS-CoV-2 variability. In the receptor binding domain of the spike protein, we observe that the predicted mutability correlates well with experimental measures of protein stability and that both are reliable mutability predictors (ROC AUC ~0.8). Most interestingly, we observe an increasing agreement between our model and the observed variability as more data become available over time, proving the anticipatory capacity of our model. When combined with data concerning the immune response, our approach identifies positions where current variants of concern are highly overrepresented. These results could assist studies on viral evolution, future viral outbreaks and, in particular, guide the exploration and anticipation of potentially harmful future SARS-CoV-2 variants.

==========

35 Denis Roze roze@sb-roscoff.fr
CNRS
Title: Recombination and the dynamics of transposable elements
Authors: Roze D

Abstract: Transposable elements (TEs) are ubiquitous among eukaryotes, constituting an important fraction of the genome of many species. TE insertion is believed to be generally neutral or deleterious, due to the disruption of gene function or regulatory sequence, or to the possibility of ectopic recombination leading to chromosomal rearrangements. This load of TEs may have important consequences on various evolutionary processes, including the evolution of recombination. Using a combination of analytical and simulation results, I will show that the transposition process generates an excess variance (positive linkage disequilibria) in the number of TEs per genome even in the presence of negative epistasis (that may stem from the possibility of ectopic recombination). While this excess variance remains small in randomly mating populations, it may become important as the rate of sex or outcrossing decreases. On the other hand, the Hill-Robertson effect tends to generate

negative LD, the sign and overall amount of LD depending on the balance between these two processes. Consequences for the evolution of recombination will be discussed.

==========

36 Stéphanie Sherpa stephanie.sherpa@hotmail.fr
Université Grenoble Alpes, LECA UMR CNRS-UGA-USMB 5553, Grenoble, France; University of Milan, Department of Environmental Science and Policy, Milano, Italy
Title: Can we predict the genetic and phenotypic traits of invasive populations of the Asian tiger mosquito in temperate and tropical regions?
Authors: Stéphanie Sherpa, Laurence Després

Abstract: The success of invasive species depends on the match between their life history traits and the environmental characteristics (biotic and abiotic) of newly colonized habitats. If the environmental characteristics between the native and the introduced ranges are similar, the success of establishment of introduced populations could be determined by pre-existing adaptation that occurred within the native range before introduction. We evaluated the role of pre- versus post-introduction adaptation in the global invasion of the Asian tiger mosquito (Aedes albopictus). In its Asian native range, A. albopictus occurs along a wide latitudinal gradient, from tropical regions of Southeast Asia to temperate regions of Japan. We studied the cold tolerance and wing-size changes along the tropical-to-temperate gradient in the native range using phenotypic traits (common garden) and SNPs in candidate genes (exon capture). We identified signatures of selection in several candidate genes where SNPs strongly correlate with temperature, including genes involved in lipid metabolism and a circadian clock gene involved in insect diapause. Allele frequencies at candidate SNPs show two abrupt shifts supporting three ecotypes. In contrast, cold tolerance and wing size exhibit phenotypic clines along the temperature gradient. Aedes albopictus conforms to Bergmann's rule and non-diapausing eggs from temperate populations survive better to cold exposure than those from tropical populations. Aedes albopictus has colonized the tropical and temperate regions of all continents. Based on ddRADseq SNPs and a large taxon sampling in East Asia (41 populations) and two replicate invasions in temperate (USA, Italy) and tropical (Cameroon, La Réunion) regions, we confirmed the existence of three evolutionary lineages within the native range, including A) Malaysia, Thailand, Cambodia, and Laos, B) China and Okinawa, and C) South Korea and Japan, and show that the niche of the invaded ranges corresponds to the climatic niche of their respective source populations. We then compared the phenotype (fitness, size, and cold tolerance) and genotype (allele frequency at outlier SNPs) of invasive populations expected based on phenotypic clines and genetic shifts in candidate genes observed in Asia (predictive maps) to the observed phenotypic and genetic traits of invasive populations.

==========

37 Yang Song y.song1@uu.nl
Utrecht University
Title: Microbiome Sequence-based tool for potato vitality prediction
Authors: Yang Song, Peter A.H.M. Bakker, Corné M.J. Pieterse, Roeland L. Berendsen

Abstract: Potato is globally one of the most important food crops and the Netherlands is leading in potato breeding and propagation. However, potato production also uses the greatest quantities of agrochemicals compared to other crops. Plant microbiomes have generated great interest as an integral part of plant biology with great new opportunities for microbiome-assisted agriculture as a biological alternative for chemical fertilizers and pesticides. Recently, we used large-scale, sequence-based microbiomics to decipher the composition of the microbiome of potato. Using bioinformatics and machine learning approaches, with microbiome and vitality data from 19 fields, we provided proof-of-concept that sequence-based potato microbiome fingerprinting of seed potato tubers can be used to predict the vitality of the potato crop that emerges from them. We will systematically

determine potato tuber compartment-specific microbiomes of seed potato tubers from 240 different Dutch seed potato fields over 2 years to link potato microbiome fingerprints to potato vitality and health and use this information to develop a microbiome sequence-based diagnostic model that is able to predict the vitality and health of seed potato batches. This model can be developed for use by seed potato breeders and growers to assess the quality of batches of seed potatoes. This will allow growers to take specific measures to improve the quality of their potatoes and improve trust between potato breeders and their potato-growing customers.

==========

38 Roman Stetsenko roman.stetsenko@sb-roscoff.fr
Station Biologique de Roscoff (Sorbonne Université / CNRS)
Title: Effect of partial self-fertilization on indirect selection on recombination
Authors: Roman Stetsenko, Denis Roze

Abstract: Recombination is thought to be one of the main advantages of sexual reproduction in eukaryotes and its evolution through the genetic associations that it generates (indirect selection) has been theoretically studied for many years, most of the work focusing on randomly mating populations. Self-fertilization can have various effects on the evolution of recombination although a positive relationship between selfing rate and recombination rate is documented in plants. The only analytical model studying the effect of partial selfing on the evolution of recombination (Roze and Lenormand, 2005) found that recombination is favored in infinite populations only when double homozygotes are less fit than expected from the fitness effect of single homozygotes (negative dominance-by-dominance epistasis). However, this three-locus model is based on approximations that break down when the selfing rate is too high and/or selected loci are tightly linked. We reanalysed this deterministic model relaxing those approximations and also studied the effect of selection and drift on the indirect selection for recombination (the Hill-Robertson effect) under partial selfing by adapting previous approximations (Roze, 2021). This three-locus model was then extended to a whole chromosome, recurrently introducing deleterious mutations and assuming a direct fitness cost of recombination. We found that selfing can either increase or decrease indirect selection for recombination depending on the parameter values. This effect of selfing is greatly reduced when a stabilizing direct selection on recombination is considered. The implications of these results concerning the expected recombination rates in partially selfing species is discussed.

==========

39 Hugo Talibart hugo.talibart@mnhn.fr
Institut Systématique Evolution Biodiversité (ISYEB), Sorbonne Université, MNHN
Title: PPalign: optimal alignment of Potts models representing proteins with direct coupling information
Authors: Hugo Talibart, François Coste

Abstract: To assign structural and functional annotations to the ever increasing amount of sequenced proteins, the main approaches rely on sequence-based homology search methods, e.g. BLAST or on profile Hidden Markov Model methods. While powerful, these approaches use positional sequence composition (unidimensional) but do not take coevolution between residues into account (3D information). Taking advantage of recent advances in the field of contact prediction, we propose to represent proteins by Potts models, which model direct couplings between positions in addition to positional composition, and to compare proteins by aligning these models. Due to non-local dependencies, the problem of aligning Potts models is hard and remains the main computational bottleneck for their use. In this talk, I will present PPalign, our method to compute the optimal pairwise alignment of Potts models representing proteins in tractable time thanks to an Integer Linear Programming formulation, which was recently published in BMC Bioinformatics. I will describe its new results in terms of alignment quality on a larger and more diversified dataset of reference pairwise

sequence alignments drawn from BALIBASE, homstrad, oxbench and sisyphus. I will show the contribution of direct couplings in improving the alignments of homologous proteins, providing insights on how homology search can be improved by taking underlying evolutional and structural constraints into account.

==========

40 Noa Vazeux-Blumental noa.vazeux-blumental@universite-paris-saclay.fr
INRAE

Title: DIVERSITY AND ORIGINS OF ASSOCIATED MAIZE AND BEANS IN ROMANIA

Authors: Vazeux-Blumental N., Manicacci D., Tenaillon M.I.

Abstract: The emergence of elite varieties used in monospecific stands is an inherent part of modern agriculture which tends to minimize interactions between species, thus opposing the natural and human selection that have been operating for millennia in a context of multi-specific assemblages. Despite the potentially interesting synergies through complementation, resource sharing, and facilitation processes, the associated cultivation of maize (Zea mays ssp. mays) and bean (Phaseolus vulgaris) imported from their Mesoamerican centre of origin has been progressively abandoned in Europe. Transylvania is one of the few regions in Europe where the intercropping of traditional varieties (landraces) of maize and bean is still practiced. These landraces therefore have the potential to be better adapted to intercropping. In order to characterize their genetic diversity and origins, we generated genotyping and phenotyping data. Our analyses revealed that associated maize and beans from Transylvania display two contrasted histories: maize used for polenta are morphologically closer to the flint-type, yet they display genetic proximity to the US Corn Belt Dents, introduced in Europe in the 19th century; beans relate to the two major American gene pools (Mesoamerican and Andean) and display no particular specificities with regards to other European beans. Our results hence do not support the hypothesis of a co-introduction of beans and maize in Transylvania. I will present the perspectives of this work which I will be exploring during my PhD.

==========

41 Lucile Vigué lucile.vigue@inserm.fr
IAME, Inserm U1137

Title: Deciphering epistasis and context-dependence contribution to polymorphism in 61,157 E. coli genomes

Authors: Lucile Vigué*, Giancarlo Croce*, Marie Petitjean, Etienne Ruppé, Olivier Tenaillon, Martin Weigt

Abstract: Characterizing the effect of mutations is key to understand the evolution of protein sequences and to separate neutral amino-acid changes from deleterious ones. Epistatic interactions between residues can lead to a context dependence of mutation effect. Context dependence constrains the amino-acid changes that can contribute to polymorphism in the short term, and the ones that can accumulate between species in the long term. We use computational approaches to accurately predict the polymorphisms segregating in a panel of 61,157 Escherichia coli genomes from the analysis of distant homologues. By comparing a context-aware Direct-Coupling Analysis modelling to a non-epistatic approach, we show that the genetic context strongly constrains the tolerable amino acids in 30% to 50% of amino-acid sites. The study of more distant species suggests the gradual build-up of genetic context over long evolutionary timescales by the accumulation of small epistatic contributions.

==========

42 Raza Waseem waokara@gmail.com
Utrecht Univeristy

Title: Ralstonia solanacearum can rapidly evolve more tolerant to volatile organic compounds produced by antagonistic bacteria
Authors: Waseem Raza, Wang Jianning, Bryden Fields, Ville-Petri Friman, Alexandre Jousset, Wei Zhong, Shen Qirong

Abstract: The soil microbes produce volatile organic compounds (VOCs) that can diffuse over long distances and have the ability to suppress pathogen and improve plant growth and resistance. However, in soil, the production of VOCs is greatly influenced by food sources and microbial community interactions. The bacterial wilt pathogen, Ralstonia solanacearum (RS), has to compete with bacteria, fungi and other microbes in the soil before infecting plants. One-way microbes can suppress RS growth is through VOCs. Here we investigated the phenotypic and genetic basis of RS adaptation to VOCs produced by Bacillus amyloliquefaciens T-5 for approximately 550 bacterial generations at three VOC levels that suppressed the RS growth by 15% (low), 30% (intermediate) and 55% (high), respectively. After the experiment, we isolated 18 RS clones from each control (no VOC) and high-level VOC treatments and sequenced their whole genomes. We found that RS growth, motility, antioxidant activity, biofilm formation and extracellular polysaccharide production were reduced in all VOC treatments. However, after 2, 4 and 6 days at low, intermediate and high levels of VOCs, respectively, RS recovered and improved growth and motility traits, while other traits decreased consistently and then kept a constant level until the end of the experiment. The genome sequence analysis showed that the clones from control and high-level VOC treatments, both showed single nucleotide polymorphism (SNP), indels and insertions compared to the ancestral RS strain. The control treatment clones showed mutations in four genes, while high-level VOC treatment showed mutations in nine genes of chromosome and five genes of megaplasmid. Mainly, in control, an intergenic mutation upstream of glycosyltransferase family 2 protein and a deletion in response regulator protein were found in 11 out of 18 clones. In high-level VOC treatment, missense SNP in glycosyltransferase (wecA) and an insertion in fimbrial type-4 assembly protein (pilM) were found in 12 out of 18 clones. Compared to ancestral strain, high-level VOC treatment clones showed increased resistance to selected VOCs and antibiotics and lost the ability to cause disease in planta. Together, these results suggest that B. amyloliquefaciens VOC selection leads to trade-offs in RS virulence and life-history traits, which indicates that biocontrol bacteria could be used to select for less virulent pathogen genotypes.

==========
43 Diego Javier Zea diegozea@gmail.com

Title: What determines the evolutionary rate of amyloid proteins?
Authors: Diego Javier Zea, Guillermo Benìtez, Juan Mac Donagh, Cristian Guisande Donadio, Nicolas Palopoli, Julia Marchetti, Maria Silvina Fornasari, Gustavo Parisi

Abstract: Background: Amyloid fibrils are insoluble, closely packed, and highly ordered arrangements that most proteins can adopt, although with different tendencies. They arise when a soluble protein or protein fragment aggregates as a filament. Filament's monomers stick between them through an array of beta-sheets perpendicular to the fiber's axis. There are plenty of studies associating amyloids with human pathologies, such as Alzheimer's and Parkinson's diseases. However, amyloids can also have functional roles in humans thanks to their high stability of the amyloid state. This study aims to shed light on the evolution of functional and pathological amyloid proteins. Results: This work studies the evolutionary rates of 65 human proteins with reported capacity to form functional or pathological amyloids compared to others. Despite being highly expressed, we have found that functional and pathological amyloids are also fast-evolving proteins. Therefore, amyloids behave differently from most proteins in evolutionary terms. We performed a multivariate analysis, including many factors affecting the evolutionary rates of proteins. That analysis shows that the main determinants of amyloids' evolutionary rates are cellular location, the number of disulfide bonds, and the interaction

with chaperones. Conclusions: These results show that most amyloids are located outside the cytoplasm, avoiding the purifying pressure acting on other highly expressed proteins to prevent cytotoxicity. At the same time, their disulfide bonds and their interactions with intra- and extra-cellular chaperones stabilize their structures, allowing them to accumulate mutations. The previously mentioned factors promote the accumulation of destabilizing mutations on amyloids, further increasing their amyloidogenic propensity.